

# Pseudo best estimator by a separable approximation of spatial covariance structures

Toshihiro Hirano\*

January 1, 2013

## Abstract

We consider the linear regression model with a spatially correlated error term on a lattice process. When we estimate coefficients in the linear regression model, the generalized least squares estimator (GLSE) is used if the covariance structures are known. However, the GLSE for large spatial data sets is impractically time-consuming because it includes the inversion of the covariance matrix of error terms in different spatial points that is the size of the number of observations. To reduce the computational complexity, we propose the pseudo best estimator (PBE) using spatial covariance structures approximated by separable covariance functions. We derive the asymptotic covariance matrix of the PBE and compare it with those of the least squares estimator (LSE) and the GLSE through some simulations. They also imply that the effect of the misspecification of the covariance matrix for the GLSE is examined. Monte Carlo simulations demonstrate the improvement of the LSE, which does not contain the information of the spatial covariance structure, by the PBE using separable covariance functions even if the true process has an isotropic Matérn covariance function. Additionally our proposed PBE is computationally efficient relative to the GLSE for large spatial data sets.

*Key words* : Asymptotic covariance matrix, Generalized least squares, Lattice process, Pseudo best estimator, Separable process, Spatial statistics, Spectral density

---

\*Graduate School of Economics, University of Tokyo, Hongo 7-3-1 Bunkyo-ku, Tokyo 113-0033 Japan. Email: 5820754252@mail.ecc.u-tokyo.ac.jp

# 1 Introduction

Recently various statistical methods for spatial data has been investigated. Among them, a linear regression model with a spatially correlated error term has been playing an important role in a wide variety of scientific fields such as geostatistics, econometrics and forestry.

To estimate coefficients in the linear regression model, we often use the generalized least squares estimator (GLSE). However, the GLSE for large spatial data sets is impractically time-consuming because it includes the inversion of the covariance matrix of error terms in different spatial points  $n$  which requires  $O(n^3)$  operations. For example, Isaaks and Srivastava (1989) and Huang et al. (2010) introduces the Walker Lake data set and the soil moisture index derived from other GIS layers which are observed on a spatial lattice of  $260 \times 300$  and  $100 \times 100$  regular grid points respectively. There is a rich literature on the analysis of large spatial data sets (Furrer et al. 2006; Banerjee et al. 2008; Cressie and Johannesson 2008; Lindgren et al. 2011).

To deal with this problem, we propose the pseudo best estimator (PBE) using spatial covariance structures approximated by the covariance function which is expressed by the product of the covariance functions of the causal autoregressive process. The covariance matrix of the PBE has the separable structure and it facilitates computational procedures for large spatial data sets by the property of the Kronecker product. Genton (2007) applied the separable approximation to the prediction of large space-time data sets by solving the nearest Kronecker product for a space-time covariance matrix problem and obtained the good predictive performance.

In this paper, we derive the asymptotic covariance matrix of the PBE for a lattice process and compare it with those of the least squares estimator (LSE) and the GLSE in Yajima and Matsuda (2008) through some simulations. The LSE is considerable efficient in terms of the fast calculation of the estimator for large spatial data sets and Yajima and Matsuda (2008) obtained the necessary and sufficient conditions where the asymptotic covariance matrix of the LSE is identical with that of the GLSE. However, our simulations illustrate that the PBE improves the accuracy of the LSE substantially when these conditions are not satisfied and shows a good performance as well as the LSE even if these conditions hold. In simulations, the difference in the mean squared error between the PBE approximated by separable covariance functions and the GLSE is small even in the case of the true process with the isotropic Matérn covariance function. Additionally, our proposed PBE is computationally efficient relative to the GLSE.

Our work can be regarded as an extension of Amemiya (1973) and Engle (1974) which considered the asymptotic properties of the GLSE when the

covariance structure of the true process was incorrectly identified in time series literature. The GLSE is infeasible unlike the PBE because the true covariance matrix is unknown in practice. Our paper also examines the effect of the misspecification of the covariance function for the GLSE in spatial statistics. Koreisha and Fang (2001) investigated the finite accuracy of the GLSE and the PBE in time series case.

The subsequent sections are organized as follows. We introduce a linear regression model with a spatially correlated error term and propose the PBE by separable covariance functions in Section 2. In Section 3, we derive the asymptotic covariance matrix of the PBE and introduce the necessary and sufficient conditions where the asymptotic covariance matrix of the LSE is identical with that of the GLSE. In Section 4, computer experiments are conducted to see the finite sample performance of our asymptotic result and compare the asymptotic variance of the PBE with that of the LSE. A conclusion and future studies are mentioned in Section 5. Technical proofs of the lemmas and the theorem are given in Appendices A and B.

## 2 Linear regression model and some estimators

From now on for simplicity we consider the sampling region of the square on the plane. Define  $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ . Let  $\mathbb{Z}^2$  be the integer lattice points in the 2-dimensional Euclidean space. For  $\mathbf{t} = (t_1, t_2)' (\in \mathbb{Z}^2)$ , consider the regression model of the form

$$y_{\mathbf{t}} = X_{\mathbf{t}}' \boldsymbol{\beta} + \epsilon_{\mathbf{t}},$$

where  $\{y_{\mathbf{t}}\}$  is an observed sequence,  $X_{\mathbf{t}} = (x_{t,1}, \dots, x_{t,p})'$  is a  $p$ -vector of nonstochastic regressors,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$  is a vector of unknown regression coefficients and the prime denotes the transposition. Hereafter it is assumed that  $(y_{\mathbf{t}}, X_{\mathbf{t}})$  is observed on the sampling domain  $P_N = \{\mathbf{t} = (t_1, t_2)' \in \mathbb{Z}^2 | 1 \leq t_1 \leq N, 1 \leq t_2 \leq N\}$ . The error terms  $\{\epsilon_{\mathbf{t}}\}$  follow a stationary random field with mean 0 and the spectral density function  $f(\boldsymbol{\lambda})$ ,  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)' \in [-\pi, \pi]^2$ . Then the covariance function  $\gamma_{\epsilon}(\mathbf{h})$  of  $\{\epsilon_{\mathbf{t}}\}$  is given by

$$\gamma_{\epsilon}(\mathbf{h}) = E[\epsilon_{\mathbf{t}} \epsilon_{\mathbf{t}+\mathbf{h}}] = \int_{\Pi^2} \exp(i\mathbf{h}'\boldsymbol{\lambda}) f(\boldsymbol{\lambda}) d\boldsymbol{\lambda},$$

where  $\Pi = (-\pi, \pi]$ ,  $\mathbf{h} = (h_1, h_2)' (\in \mathbb{Z}^2)$  and  $\mathbf{h}'\boldsymbol{\lambda} = h_1\lambda_1 + h_2\lambda_2$ .

When we estimate coefficients in the linear regression model, the GLSE is given by

$$\hat{\boldsymbol{\beta}}_{GLSE} = \left( X' \Sigma^{-1} X \right)^{-1} X' \Sigma^{-1} \mathbf{y},$$

where  $X = (X_{(1,1)}, \dots, X_{(1,N)}, X_{(2,1)}, \dots, X_{(2,N)}, \dots, X_{(N,1)}, \dots, X_{(N,N)})'$ ,  $\mathbf{y} = (y_{(1,1)}, \dots, y_{(1,N)}, y_{(2,1)}, \dots, y_{(2,N)}, \dots, y_{(N,1)}, \dots, y_{(N,N)})'$ ,  $\Sigma = E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}']$  and  $\boldsymbol{\epsilon} = (\epsilon_{(1,1)}, \dots, \epsilon_{(1,N)}, \epsilon_{(2,1)}, \dots, \epsilon_{(2,N)}, \dots, \epsilon_{(N,1)}, \dots, \epsilon_{(N,N)})'$ .

It is known that the GLSE is the best linear unbiased estimator (BLUE). However, if the true covariance structure is unknown, the GLSE is infeasible. Additionally, the operation count for computing  $\Sigma^{-1}$  ( $N^2 \times N^2$ ) of  $\hat{\boldsymbol{\beta}}_{GLSE} \left( = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} \mathbf{y} \right)$  is of order  $N^6$ . Hence as the sample size is larger, the computation becomes a more formidable one in practice. For example, the Walker Lake data set in Isaaks and Srivastava (1989) consists of two variables measured at 78000 points on a spatial lattice of  $260 \times 300$  regular grid points. In Huang et al. (2010), there is the soil moisture index derived from other GIS layers on  $100 \times 100$  points.

To reduce the computational burden, we consider the approximation of the true covariance function by the product of the covariance functions of the causal autoregressive process of order  $P$  (AR( $P$ )) in time series literature. Then we obtain the following estimator

$$\hat{\boldsymbol{\beta}}_{PBE} = \left( X' \tilde{\Sigma}^{-1} X \right)^{-1} X' \tilde{\Sigma}^{-1} \mathbf{y},$$

where  $\tilde{\Sigma} = \tilde{\Sigma}_1 \otimes \tilde{\Sigma}_2$  and  $\tilde{\Sigma}_i$  is the covariance matrix of causal AR( $P_i$ ) ( $i = 1, 2$ ). This kind of estimator, in which the true covariance matrix is replaced with the incorrect one, is called the pseudo best estimator. Each element of  $\tilde{\Sigma}$  is denoted by the separable covariance function  $\gamma(h_1, h_2) = \gamma_1(h_1) \times \gamma_2(h_2)$  and  $\gamma_i(h_i)$  represents the autocovariance function of AR( $P_i$ ) ( $i = 1, 2$ ). Note that  $h_1$  and  $h_2$  of  $\gamma(h_1, h_2)$  correspond to those of  $\gamma_\epsilon(h_1, h_2)$ . From the property of the Kronecker product,  $\tilde{\Sigma}^{-1} = \left( \tilde{\Sigma}_1 \otimes \tilde{\Sigma}_2 \right)^{-1} = \tilde{\Sigma}_1^{-1} \otimes \tilde{\Sigma}_2^{-1}$  and we can obtain the exact form of the inverse matrix of the covariance one  $\tilde{\Sigma}_i$  ( $i = 1, 2$ ) of the autoregressive process (e.g., Anderson 1974; page 576), so that it is much faster to calculate  $\hat{\boldsymbol{\beta}}_{PBE}$  by the separable approximation of the true covariance function.

The LSE is also the alternative candidate for the GLSE because it does not include the inversion of the covariance matrix and is defined by

$$\hat{\boldsymbol{\beta}}_{LSE} = \left( X' X \right)^{-1} X' \mathbf{y}.$$

We will compare the accuracy of the GLSE, the PBE and the LSE by evaluating the asymptotic covariance matrix.

### 3 The asymptotic properties of $\hat{\beta}_{GLSE}$ , $\hat{\beta}_{LSE}$ and $\hat{\beta}_{PBE}$

In this section, the asymptotic covariance matrices of  $\hat{\beta}_{GLSE}$ ,  $\hat{\beta}_{LSE}$  and  $\hat{\beta}_{PBE}$  are derived. First we introduce some assumptions.

Define

$$\begin{aligned} a_{ij}^{(N,N)}(h_1, h_2) &= \sum_{t_1=1}^{N-h_1} \sum_{t_2=1}^{N-h_2} x_{(t_1+h_1, t_2+h_2), i} x_{(t_1, t_2), j}, \quad h_1, h_2 \geq 0, \\ &= \sum_{t_1=1-h_1}^N \sum_{t_2=1-h_2}^N x_{(t_1+h_1, t_2+h_2), i} x_{(t_1, t_2), j}, \quad h_1, h_2 \leq 0, \\ &= \sum_{t_1=1-h_1}^N \sum_{t_2=1}^{N-h_2} x_{(t_1+h_1, t_2+h_2), i} x_{(t_1, t_2), j}, \quad h_1 \leq 0, h_2 \geq 0, \\ &= \sum_{t_1=1}^{N-h_1} \sum_{t_2=1-h_2}^N x_{(t_1+h_1, t_2+h_2), i} x_{(t_1, t_2), j}, \quad h_1 \geq 0, h_2 \leq 0. \end{aligned}$$

- (a)  $a_{ii}^{(N,N)}(0, 0) \rightarrow \infty$  as  $N \rightarrow \infty, i = 1, \dots, p$ .
- (b)  $\lim_{N \rightarrow \infty} a_{ii}^{(N+h_1, N+h_2)}(0, 0) / a_{ii}^{(N,N)}(0, 0) = 1$  for every  $i$  and  $h_1, h_2, i = 1, \dots, p$  and  $h_1, h_2 \in \mathbb{Z}$ .
- (c) The limit of

$$\gamma_{ij}^{(N,N)}(h_1, h_2) = \frac{a_{ij}^{(N,N)}(h_1, h_2)}{\{a_{ii}^{(N,N)}(0, 0) a_{jj}^{(N,N)}(0, 0)\}^{1/2}}$$

as  $N \rightarrow \infty$  exists for every  $i, j$  and  $h_1, h_2, (i, j = 1, \dots, p$  and  $h_1, h_2 \in \mathbb{Z})$ .

Let

$$\rho_{ij}(h_1, h_2) = \lim_{N \rightarrow \infty} \gamma_{ij}^{(N,N)}(h_1, h_2)$$

and  $R(h_1, h_2)$  be the  $p \times p$  matrix with  $(i, j)$ th element  $\rho_{ij}(h_1, h_2)$ .

- (d)  $R(0, 0)$  is nonsingular.
- (e)  $\{\epsilon_t\}$  is a unilateral MA process,

$$\epsilon_{(t_1, t_2)} = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \theta_{l, m} \eta_{(t_1-l, t_2-m)},$$

where  $\{\theta_{l, m}\}$  satisfies  $\sum_{l, m} |\theta_{l, m}|^2 < \infty$  with  $\theta_{0, 0} = 1$  and  $\{\eta_{(t_1, t_2)}\}$  is a white noise with  $Var(\eta_{(t_1, t_2)}) = \sigma_{\eta}^2$ .

Moreover if we define  $\theta(z_1, z_2) = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \theta_{l,m} z_1^l z_2^m$  and  $\phi(z_1, z_2) = \theta(z_1, z_2)^{-1} = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \phi_{l,m} z_1^l z_2^m$ , then  $\{\phi_{l,m}\}$  satisfies  $\sum_{l,m} |\phi_{l,m}| < \infty$  with  $\phi_{0,0} = 1$ .

(f)  $f(\boldsymbol{\lambda})$  is a positive continuous function in  $[-\pi, \pi]^2$ .

(g)  $N \max_{1 \leq t_1, t_2 \leq N} (x_{(t_1, t_2), i})^2 / a_{ii}^{(N, N)}(0, 0) \rightarrow 0$  as  $N \rightarrow \infty$  ( $i = 1, \dots, p$ ).

(h)  $f(\lambda_1, \lambda_2) = f(-\lambda_1, \lambda_2) = f(\lambda_1, -\lambda_2) = f(-\lambda_1, -\lambda_2)$ .

We make brief comments on the assumptions. We can view (a)-(d) as a two-dimensional version of so-called Grenander's conditions on  $X_t$ . Under (a)-(d), there exists a Hermitian matrix function  $M(\lambda_1, \lambda_2)$  with positive semidefinite increments such that

$$R(h_1, h_2) = \int_{\Pi^2} \exp(i(h_1 \lambda_1 + h_2 \lambda_2)) dM(\lambda_1, \lambda_2)$$

(see Cohen and Francos 2002). Put

$$m_{kl}^{(N, N)}(\lambda_1, \lambda_2) = \left( \sum_{t_1=1}^N \sum_{t_2=1}^N x_{(t_1, t_2), k} e^{-i(t_1 \lambda_1 + t_2 \lambda_2)} \right) \left( \sum_{t_1=1}^N \sum_{t_2=1}^N x_{(t_1, t_2), l} e^{i(t_1 \lambda_1 + t_2 \lambda_2)} \right) \\ \Bigg/ \left( (2\pi)^2 \left( a_{kk}^{(N, N)}(0, 0) a_{ll}^{(N, N)}(0, 0) \right)^{1/2} \right)$$

and define

$$M_{kl}^{(N, N)}(\lambda_1, \lambda_2) = \int_{-\pi}^{\lambda_1} \int_{-\pi}^{\lambda_2} m_{kl}^{(N, N)}(\omega_1, \omega_2) d\omega_1 d\omega_2.$$

Let  $M_{kl}(\lambda_1, \lambda_2)$  and  $M_{kl}^{(N, N)}(\lambda_1, \lambda_2)$  be  $p \times p$  matrices with  $(k, l)$ th element of  $M(\lambda_1, \lambda_2)$  and  $M^{(N, N)}(\lambda_1, \lambda_2)$  respectively for  $k, l = 1, \dots, p$ . If we regard  $M^{(N, N)}(\lambda_1, \lambda_2)$  and  $M(\lambda_1, \lambda_2)$  as matrix measures in  $\Pi^2$ ,  $R^{(N, N)}(h_1, h_2) = [\gamma_{kl}^{(N, N)}(h_1, h_2)]$  and  $R(h_1, h_2)$  are their characteristic functions respectively. Then (c) implies

$$M^{(N, N)}(\lambda_1, \lambda_2) \xrightarrow{w} M(\lambda_1, \lambda_2)$$

as  $N \rightarrow \infty$ , where  $\xrightarrow{w}$  means  $M^{(N, N)}(\lambda_1, \lambda_2)$  converges weakly to  $M(\lambda_1, \lambda_2)$ . Consequently, for any continuous bounded function  $\phi(\boldsymbol{\lambda})$  in  $\Pi^2$ ,

$$\lim_{N \rightarrow \infty} \int_{\Pi^2} \phi(\boldsymbol{\lambda}) dM^{(N, N)}(\lambda_1, \lambda_2) = \int_{\Pi^2} \phi(\boldsymbol{\lambda}) dM(\lambda_1, \lambda_2).$$

$M(\lambda_1, \lambda_2)$  is called the regression spectral measure of  $X_t$  in accordance with the time series literature (Taniguchi et al. 2008).

(e) is somewhat restrictive, but (e) holds for a random process with the separable covariance functions such as the product of 1-dimensional Matérn

covariance functions (Yajima and Matsuda 2008). (h) is often called axial symmetry. The spectral density functions of a separable covariance and isotropic one satisfy (h).

Next we state the asymptotic covariance matrix of three estimators, that is  $\hat{\beta}_{GLSE}$ ,  $\hat{\beta}_{PBE}$  and  $\hat{\beta}_{LSE}$ .

Define

$$D_{N^2} = \text{diag}(\|\mathbf{x}_1\|, \dots, \|\mathbf{x}_p\|)$$

where  $\mathbf{x}_i = (x_{(1,1),i}, \dots, x_{(1,N),i}, x_{(2,1),i}, \dots, x_{(2,N),i}, \dots, x_{(N,1),i}, \dots, x_{(N,N),i})'$  and  $\|\mathbf{x}_i\| = (\sum_{t_1=1}^N \sum_{t_2=1}^N (x_{(t_1,t_2),i})^2)^{1/2}$  ( $i = 1, \dots, p$ ).

**Theorem 1 (Yajima and Matsuda (2008)).** Under (a)-(g),

$$\begin{aligned} & \lim_{N \rightarrow \infty} D_{N^2} E \left[ (\hat{\beta}_{GLSE} - \beta)(\hat{\beta}_{GLSE} - \beta)' \right] D_{N^2} \\ &= \left( \frac{1}{(2\pi)^2} \int_{\Pi^2} \frac{1}{f(\lambda_1, \lambda_2)} dM(\lambda_1, \lambda_2) \right)^{-1}. \end{aligned}$$

**Theorem 2 (Yajima and Matsuda (2008)).** Under (a)-(d) and (f),

$$\begin{aligned} & \lim_{N \rightarrow \infty} D_{N^2} E \left[ (\hat{\beta}_{LSE} - \beta)(\hat{\beta}_{LSE} - \beta)' \right] D_{N^2} \\ &= (2\pi)^2 R(0, 0)^{-1} \int_{\Pi^2} f(\lambda_1, \lambda_2) dM(\lambda_1, \lambda_2) R(0, 0)^{-1}. \end{aligned}$$

Moreover, Yajima and Matsuda (2008) gives the following necessary and sufficient conditions where the asymptotic covariance matrix of the LSE is identical with that of the GLSE.

**Theorem 3 (Yajima and Matsuda (2008)).** Under (a)-(g), the LSE is asymptotically efficient relative to the GLSE (BLUE) if and only if  $M(\lambda_1, \lambda_2)$  increases at not more than  $p$  values of  $\lambda$  ( $\in [0, \pi]^2$ ) and the sum of the ranks of the increases in  $M(\lambda_1, \lambda_2)$  is  $p$ .

Theorem 1-3 can be regarded as an extension of Grenander and Rosenblatt (1957) for  $d = 1$  to spatial processes. The following theorem is our main result. The proof is given in Appendix B.

**Theorem 4.** Under (a)-(d) and (f)-(h),

$$\begin{aligned} & \lim_{N \rightarrow \infty} D_{N^2} E \left[ (\hat{\beta}_{PBE} - \beta)(\hat{\beta}_{PBE} - \beta)' \right] D_{N^2} \\ &= (2\pi)^2 \left( \int_{\Pi^2} \frac{1}{g(\lambda_1, \lambda_2)} dM(\lambda_1, \lambda_2) \right)^{-1} \left( \int_{\Pi^2} \frac{f(\lambda_1, \lambda_2)}{g^2(\lambda_1, \lambda_2)} dM(\lambda_1, \lambda_2) \right) \\ &\times \left( \int_{\Pi^2} \frac{1}{g(\lambda_1, \lambda_2)} dM(\lambda_1, \lambda_2) \right)^{-1}. \end{aligned}$$

Note that conditions of Theorems 1-4 are somewhat different. In particular, unlike Theorems 2 and 4, Theorems 1 and 3 do not hold due to (e) if the random process has the isotropic Matérn covariance function. The separable covariance functions can satisfy all the conditions of Theorems 1-4.

Amemiya (1973) and Engle (1974) investigated the asymptotic properties of  $\hat{\beta}_{PBE}$  for  $d = 1$  case and Theorem 4 is an extension of their results to the spatial case.

## 4 Computational experiments

We conduct Monte Carlo simulation by using MATLAB. The finite accuracy and the asymptotic variance of the PBE are compared with those of the LSE through computational simulations especially in the case where the LSE is not asymptotically efficient. We consider the linear regression models with one regressor

$$y_{(t_1, t_2)} = \beta x_{(t_1, t_2)} + \epsilon_{(t_1, t_2)}$$

for  $1 \leq t_1, t_2 \leq N$  and  $\beta = 2$ . For the regressor  $x_{(t_1, t_2)}$ , the polynomial trend  $x_{(t_1, t_2)} = t_1 t_2$ , the harmonic trend  $x_{(t_1, t_2)} = \cos((\pi/2)t_1) \cos((\pi/2)t_2)$  and the polynomial plus harmonic trend  $x_{(t_1, t_2)} = 1 + \cos((\pi/2)t_1) \cos((\pi/2)t_2)$  (see Toyooka 1985 for  $d = 1$  case) are considered. These regressors satisfy (a)-(d) and (g). The jumps of  $M(\lambda_1, \lambda_2)$  are 1 at  $(\lambda_1, \lambda_2) = (0, 0)$  and  $1/4$  at  $(\lambda_1, \lambda_2) = (\pi/2, \pi/2), (-\pi/2, \pi/2), (\pi/2, -\pi/2), (-\pi/2, -\pi/2)$  for the polynomial trend and the harmonic trend respectively. Therefore, for these two regressors the LSE is asymptotically efficient from Theorem 3 if (e) and (f) hold. Additionally, it follows from a routine calculation that the asymptotic variance of the PBE is identical with that of the LSE in just these cases. As a result, the PBE is also asymptotically efficient under the above situation. In the third regressor, the jumps of  $M(\lambda_1, \lambda_2)$  are  $4/5$  at  $(\lambda_1, \lambda_2) = (0, 0)$  and  $1/20$  at  $(\lambda_1, \lambda_2) = (\pi/2, \pi/2), (-\pi/2, \pi/2), (\pi/2, -\pi/2), (-\pi/2, -\pi/2)$ . Since the condition of Theorem 3 is not satisfied, the LSE is not asymptotically efficient.



For the true covariance function  $\gamma_\epsilon(\mathbf{h})$ , we consider the following six models. The first and second models are

$$\gamma_\epsilon(\mathbf{h}) = c(\|\mathbf{h}\|) \quad \|\mathbf{h}\| = \sqrt{h_1^2 + h_2^2},$$

where

$$c(x) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left( \frac{2\sqrt{\nu}|x|}{\rho} \right)^\nu K_\nu \left( \frac{2\sqrt{\nu}|x|}{\rho} \right)$$

and  $K_\nu(\cdot)$  is the modified Bessel function of the second kind of order  $\nu$ . This is the isotropic Matérn covariance function and we put  $\nu = 2, \rho = 3, \sigma^2 = 1$  and  $\nu = 1, \rho = 3, \sigma^2 = 1$  respectively. The third model is the product of 1-dimensional Matérn covariance functions with  $\nu = 2, \rho = 3, \sigma^2 = 1$  and  $\nu = 1, \rho = 3, \sigma^2 = 1$  respectively. The fourth one is the product of 1-dimensional Matérn covariance function with  $\nu = 1, \rho = 3, \sigma^2 = 1$  and the autocovariance function of AR(2)

$$c^*(x) = \frac{\sigma_*^2 \xi_1^2 \xi_2^2}{(\xi_1 \xi_2 - 1)(\xi_2 - \xi_1)} \left[ \frac{\xi_1^{1-|x|}}{\xi_1^2 - 1} - \frac{\xi_2^{1-|x|}}{\xi_2^2 - 1} \right],$$

where  $\xi_1 = (2/3)(1 + \sqrt{3}i), \xi_2 = (2/3)(1 - \sqrt{3}i)$  and  $\sigma_*^2$  chosen such as  $c^*(0) = 1$ . It is inappropriate to approximate this autocovariance function  $c^*(x)$  by that of AR(1) because the successive negative correlation exists. The fifth one is the product of the autocovariance function of AR(2) in the fourth one and that of AR(1)

$$c^{**}(x) = \frac{\sigma_{**}^2}{1 - \phi^2} \phi^{|x|},$$

where  $\phi = 0.5$  and  $\sigma_{**}^2$  chosen such as  $c^{**}(0) = 1$ . The sixth one is the product of two autocovariance functions of AR(1) with  $\phi = 0.9$  and the scale parameter such as  $\gamma_\epsilon(0) = 1$ . The covariance functions except the first and second ones satisfy (e) (Yajima and Matsuda 2008) and all the models satisfy (f) and (h).

Next we explain the separable approximation like the Yule-Walker estimator in this simulation. Three types, that is AR(1)×AR(1), AR(1)×AR(2) and AR(2)×AR(2), are adopted. The first approximation is expressed by

$$\begin{aligned} \gamma(h_1, h_2) &= \gamma_1(h_1)\gamma_2(h_2) \\ &= \frac{\sigma_1^2}{1 - \phi_1^2} \phi_1^{|h_1|} \frac{\sigma_2^2}{1 - \phi_2^2} \phi_2^{|h_2|}. \end{aligned}$$

Each parameter is estimated by

$$\hat{\phi}_1 = \frac{\hat{\gamma}(1, 0)}{\hat{\gamma}(0, 0)}, \quad \hat{\phi}_2 = \frac{\hat{\gamma}(0, 1)}{\hat{\gamma}(0, 0)} \text{ and } \hat{\sigma}_{12}^2 = \frac{\hat{\gamma}(0, 0)}{\gamma' \left( 0, 0; \hat{\phi}_1, \hat{\phi}_2 \right)},$$

where

$$\begin{aligned} \hat{\gamma}(\mathbf{h}) &= \frac{1}{N(\mathbf{h})} \sum_{(i,j) \in S(\mathbf{h})} (\hat{\epsilon}_{t_i} - \bar{\hat{\epsilon}})(\hat{\epsilon}_{t_j} - \bar{\hat{\epsilon}}), \\ S(\mathbf{h}) &= \{(i, j) | \mathbf{h} = \mathbf{t}_i - \mathbf{t}_j, \mathbf{t}_i, \mathbf{t}_j \in P_N\}, \quad N(\mathbf{h}) = \#\{S(\mathbf{h})\}, \\ \hat{\epsilon}_{(t_1, t_2)} &= y_{(t_1, t_2)} - \hat{\beta}_{LSE} x_{(t_1, t_2)}, \quad \bar{\hat{\epsilon}} = \frac{1}{N^2} \sum_{t_1=1}^N \sum_{t_2=1}^N \hat{\epsilon}_{(t_1, t_2)} \end{aligned}$$

and

$$\gamma' \left( 0, 0; \hat{\phi}_1, \hat{\phi}_2 \right) = \frac{1}{(1 - \hat{\phi}_1^2)(1 - \hat{\phi}_2^2)}$$

(see Cressie 1993).  $\hat{\sigma}_{12}^2$  is the estimator of  $\sigma_1^2 \sigma_2^2$  and is not necessary for the calculation of  $\hat{\beta}_{PBE}$ . It is used in the estimation of the spectral density function  $g(\lambda_1, \lambda_2)$  as we discuss later. For the second one, the covariance function used in the separable approximation and the estimator are

$$\begin{aligned} \gamma(h_1, h_2) &= \gamma_1(h_1) \gamma_2(h_2) \\ &= \frac{\sigma_1^2}{1 - \phi_1^2} \phi_1^{|h_1|} \frac{\sigma_2^2 \xi_1^2 \xi_2^2}{(\xi_1 \xi_2 - 1)(\xi_2 - \xi_1)} \left[ \frac{\xi_1^{1-|h_2|}}{\xi_1^2 - 1} - \frac{\xi_2^{1-|h_2|}}{\xi_2^2 - 1} \right], \\ \hat{\xi}_1 &= \frac{\hat{a} + \sqrt{\hat{a}^2 + 4\hat{b}}}{-2\hat{b}} \text{ and } \hat{\xi}_2 = \frac{\hat{a} - \sqrt{\hat{a}^2 + 4\hat{b}}}{-2\hat{b}}, \end{aligned}$$

where

$$\begin{aligned} \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} &= \begin{pmatrix} \hat{\rho}_2(0) & \hat{\rho}_2(-1) \\ \hat{\rho}_2(1) & \hat{\rho}_2(0) \end{pmatrix}^{-1} \begin{pmatrix} \hat{\rho}_2(1) \\ \hat{\rho}_2(2) \end{pmatrix}, \\ \hat{\rho}_2(0) &= 1, \quad \hat{\rho}_2(1) = \frac{\hat{\gamma}(0, 1)}{\hat{\gamma}(0, 0)} = \hat{\rho}_2(-1) \text{ and } \hat{\rho}_2(2) = \frac{\hat{\gamma}(0, 2)}{\hat{\gamma}(0, 0)}. \end{aligned}$$

$\phi_1$  and  $\sigma_1^2 \sigma_2^2$  are estimated by the similar way in the case of the first approximation model. As before, the parameters are estimated in  $\text{AR}(2) \times \text{AR}(2)$  case.

Table 1: Simulation results in the polynomial trend case.

	case	Matérn ( $\nu = 2.0$ )		Matérn ( $\nu = 1.0$ )		$(\nu = 2.0) \times (\nu = 1.0)$	
approximation	N	LSE	PBE	LSE	PBE	LSE	PBE
AR(1) $\times$ AR(1)	20	1.328	1.130	1.214	1.027	1.187	1.154
	60	1.108	1.123	1.137	1.022	1.066	1.109
AR(1) $\times$ AR(2)	20	-	1.073	-	1.032	-	1.134
	60	-	1.057	-	1.021	-	1.082
AR(2) $\times$ AR(2)	20	-	1.038	-	1.031	-	1.026
	60	-	1.018	-	1.025	-	0.999
	case	$(\nu = 1.0) \times \text{AR}(2)$		AR(1) $\times$ AR(2)		AR(1) $\times$ AR(1)	
approximation	N	LSE	PBE	LSE	PBE	LSE	PBE
AR(1) $\times$ AR(1)	20	1.145	1.199	1.066	1.179	1.929	1.085
	60	1.073	1.102	1.053	1.091	1.379	0.999
AR(1) $\times$ AR(2)	20	-	1.024	-	1.006	-	1.156
	60	-	1.009	-	1.0	-	1.013
AR(2) $\times$ AR(2)	20	-	1.007	-	1.016	-	1.182
	60	-	1.003	-	1.001	-	1.019

We generate the error terms  $\{\epsilon_{(t_1, t_2)}\}_{1 \leq t_1, t_2 \leq N}$  by the multivariate normal distribution with the mean  $\mathbf{0}$  and the covariance matrix  $\Sigma$ .  $\hat{\beta}_{PBE}$  is calculated by using the parameters estimated from  $\hat{\epsilon}_{(t_1, t_2)} = y_{(t_1, t_2)} - \hat{\beta}_{LSE} x_{(t_1, t_2)}$ . The sample size is  $N^2 = 20^2$  or  $60^2$  with 1000 replications. From them, we calculate the empirical ratios of the variances of the LSE and the PBE to that of the GLSE.

Tables 1 and 2 summarize the results in the case of the first and second regressors. As  $N$  increases, the finite efficiency seems to go to 1 in all the cases. The PBE shows good performance even in the case of the isotropic Matérn class. If AR(1)  $\times$  AR(2) and AR(1)  $\times$  AR(1) are the true covariance function, the PBE by the corresponding approximation is superior to those in other cases. Moreover, although the asymptotic variances of the LSE and the PBE are equal to each other, the PBE outperforms the LSE in many cases of  $N = 60$ . That is because the PBE includes the information of the approximated spatial correlation structure. Throughout these simulations, it seems that as the order of the autoregressive process in the separable approximation is larger, the efficiency tends to be often better.

Table 3 shows the result in the polynomial plus harmonic trend case.  $(\cdot)$  means the theoretical ratios of the asymptotic variances of the LSE and the PBE in Theorems 2 and 4 to that of the GLSE in Theorem 1.  $g(\lambda_1, \lambda_2)$  with the average of values of parameters estimated in each iteration in  $N = 60$

Table 2: Simulation results in the harmonic trend case.

	case	Matérn ( $\nu = 2.0$ )		Matérn ( $\nu = 1.0$ )		$(\nu = 2.0) \times (\nu = 1.0)$	
approximation	N	LSE	PBE	LSE	PBE	LSE	PBE
AR(1) $\times$ AR(1)	20	1.178	1.033	1.040	1.025	1.483	1.049
	60	1.064	1.008	1.032	1.003	1.151	1.015
AR(1) $\times$ AR(2)	20	-	1.035	-	1.038	-	1.044
	60	-	1.012	-	1.001	-	1.017
AR(2) $\times$ AR(2)	20	-	1.045	-	1.048	-	1.024
	60	-	1.012	-	1.003	-	1.003
	case	$(\nu = 1.0) \times \text{AR}(2)$		AR(1) $\times$ AR(2)		AR(1) $\times$ AR(1)	
approximation	N	LSE	PBE	LSE	PBE	LSE	PBE
AR(1) $\times$ AR(1)	20	1.343	1.106	1.203	1.078	1.770	1.010
	60	1.119	1.033	1.075	1.032	1.271	1.0
AR(1) $\times$ AR(2)	20	-	1.017	-	1.006	-	1.016
	60	-	1.003	-	1.0	-	1.0
AR(2) $\times$ AR(2)	20	-	1.010	-	1.011	-	1.012
	60	-	1.0	-	1.0	-	1.003

Table 3: Simulation results in the polynomial plus harmonic trend case.

	case	Matérn ( $\nu = 2.0$ )		Matérn ( $\nu = 1.0$ )		$(\nu = 2.0) \times (\nu = 1.0)$	
approximation	N	LSE	PBE	LSE	PBE	LSE	PBE
AR(1) $\times$ AR(1)	20	37.347	1.024	18.060	1.069	54.357	1.051
	60	42.525 (44.431)*	1.036 (1.012)*	20.083 (21.069)*	1.022 (1.019)*	65.797 (70.077)	1.028 (1.004)
AR(1) $\times$ AR(2)	20	-	1.066	-	1.070	-	1.076
	60	-	1.036 (1.01)*	-	1.029 (1.017)*	-	1.016 (1.003)
AR(2) $\times$ AR(2)	20	-	1.085	-	1.099	-	1.105
	60	-	1.011 (1.006)*	-	1.002 (1.01)*	-	1.006 (1.002)
	case	$(\nu = 1.0) \times \text{AR}(2)$		AR(1) $\times$ AR(2)		AR(1) $\times$ AR(1)	
approximation	N	LSE	PBE	LSE	PBE	LSE	PBE
AR(1) $\times$ AR(1)	20	3.263	1.260	1.661	1.364	1691	16.18
	60	3.652 (3.575)	1.137 (1.147)	1.660 (1.622)	1.327 (1.283)	3732 (5242)	1.02 (1.0)
AR(1) $\times$ AR(2)	20	-	1.023	-	1.049	-	32.515
	60	-	0.996 (1.002)	-	1.005 (1.0)	-	1.022 (1.0)
AR(2) $\times$ AR(2)	20	-	1.068	-	1.069	-	68.844
	60	-	1.002 (1.001)	-	1.016 (1.002)	-	1.03 (1.001)

Table 4: The time required for the calculation of each estimator for three approximations (seconds).

	LSE	GLSE	PBE(AR(1)×AR(1))	PBE(AR(1)×AR(2))	PBE(AR(2)×AR(2))
estimation procedure (sec.)	-	-	0.0096	0.0086	0.013
calculation time of estimators (sec.)	$4.43 * 10^{-4}$	384.793	2.368	2.404	2.399

case is used when we calculate the asymptotic variance of  $\hat{\beta}_{PBE}$  in Theorem 4. Since the isotropic Matérn case does not satisfy (e) and the asymptotic variance of the GLSE in this case cannot be calculated by Theorem 1,  $(\cdot)^*$  means the ratios of the asymptotic variances of the LSE and the PBE in Theorems 2 and 4 to that of the GLSE calculated formally in Theorem 1. Note that the spectral density function of the Matérn covariance function must be calculated in consideration of the aliasing phenomenon (Fuentes 2005) because the observations are in the integer lattice. Since the LSE is not asymptotically efficient in the third regressor case, the performance of the LSE is too bad. However, the efficiency of the PBE is near 1 in both the empirical and theoretical ratios. As  $N$  increases, the finite efficiency seems to go to the value of  $(\cdot)$  and  $(\cdot)^*$ . This supports the theoretical results and suggests that (e) may be weakened for Theorem 1. Other properties are similar to the preceding simulations.

Next we see one realization of the spectral density function  $g(\lambda_1, \lambda_2)$  by using the average of values of parameters estimated in each iteration for the polynomial plus harmonic trend case. Figure 1 shows the true spectral density function  $f$  and the approximated spectral density functions  $g$  on  $[0, \pi]^2$  because  $f$  and  $g$  are axial symmetry. In the case of the separable approximation  $AR(2) \times AR(2)$ , except the true model of  $AR(1) \times AR(1)$ , the fitting of the spectral density function  $g$  for  $f$  is very good at the points  $(\lambda_1, \lambda_2)$  where  $M(\lambda_1, \lambda_2)$  has jumps. It is consistent with the result of Table 3 which shows the good performance of the separable approximation  $AR(2) \times AR(2)$ .

Finally we examine the computational time of  $\hat{\beta}_{GLSE}$ ,  $\hat{\beta}_{LSE}$  and  $\hat{\beta}_{PBE}$ . We put  $N = 100$  and adopt the Matérn covariance function with  $\nu = 2, \rho = 3, \sigma^2 = 1$  for the polynomial plus harmonic trend. All computations are carried out on Linux powered 3.33GHz Xeon processor with 8 Gbytes RAM. From Table 4, the LSE is computationally efficient, but Table 3 shows too bad performance in this case. The computation time of the PBE including the estimation procedure is faster than that of the GLSE and we observe the good performance of the PBE in Table 3.

## 5 Conclusion and future studies

Our study proposes the PBE by using the separable approximation of the covariance function on a lattice data as an alternative estimator of the GLSE which is infeasible due to the computational burden and the unknown covariance structure in practice. We derive the asymptotic covariance matrix of the PBE and see the effect of the misspecification of the covariance function for the the GLSE. The PBE by the separable approximation works well in many simulations even if the true covariance function is the isotropic one. In particular, when the LSE is not asymptotically efficient, the PBE shows a better performance than that of the LSE. Moreover, the PBE reduces the computational time substantially because of the separable structure.

The theoretical comparison of the asymptotic covariance matrix of the PBE and that of the LSE is a future work. In addition, the extension of the true process to the strongly dependent random field should be considered.

### Appendix A : Technical Lemmas

To derive the asymptotic covariance matrix of  $\hat{\beta}_{PBE}$ , we shall prove the two lemmas.

**Lemma A. 1.** *Let  $(\pi_{1,1}^{m_1,m_2}, \pi_{1,2}^{m_1,m_2}, \dots, \pi_{1,N}^{m_1,m_2}, \pi_{2,1}^{m_1,m_2}, \dots, \pi_{N,1}^{m_1,m_2}, \dots, \pi_{N,N}^{m_1,m_2})'$  be the solution of*

$$\tilde{\Sigma} \begin{pmatrix} \pi_{1,1}^{m_1,m_2} \\ \vdots \\ \pi_{N,N}^{m_1,m_2} \end{pmatrix} = \begin{pmatrix} \gamma(m_1, m_2) \\ \gamma(m_1, m_2 + 1) \\ \vdots \\ \gamma(m_1, m_2 + N - 1) \\ \gamma(m_1 + 1, m_2) \\ \vdots \\ \gamma(m_1 + N - 1, m_2 + N - 1) \end{pmatrix}, \quad m_1, m_2 \in \mathbb{Z}. \quad (\text{A.1})$$

For fixed  $m_1 \geq -N + 1$  and  $m_2 \geq -N + 1$ ,

$$\sup_N \sum_{i=1}^N \sum_{j=1}^N |\pi_{i,j}^{m_1,m_2}| < \infty.$$

*Proof.* (A.1) is expressed by

$$(\tilde{\Sigma}_1 \otimes \tilde{\Sigma}_2) \begin{pmatrix} \pi_{1,1}^{m_1,m_2} \\ \vdots \\ \pi_{N,N}^{m_1,m_2} \end{pmatrix} = \begin{pmatrix} \gamma_1(m_1) \\ \vdots \\ \gamma_1(m_1 + N - 1) \end{pmatrix} \otimes \begin{pmatrix} \gamma_2(m_2) \\ \vdots \\ \gamma_2(m_2 + N - 1) \end{pmatrix}.$$

Then, from the property of the Kronecker product (see Horn and Johnson 1991),

$$\begin{pmatrix} \pi_{1,1}^{m_1,m_2} \\ \vdots \\ \pi_{N,N}^{m_1,m_2} \end{pmatrix} = \left( \tilde{\Sigma}_1^{-1} \begin{pmatrix} \gamma_1(m_1) \\ \vdots \\ \gamma_1(m_1 + N - 1) \end{pmatrix} \right) \otimes \left( \tilde{\Sigma}_2^{-1} \begin{pmatrix} \gamma_2(m_2) \\ \vdots \\ \gamma_2(m_2 + N - 1) \end{pmatrix} \right).$$

Now let  $(\pi_{1,k}^{m_k}, \pi_{2,k}^{m_k}, \dots, \pi_{N,k}^{m_k})'$  be the solution of

$$\tilde{\Sigma}_k \begin{pmatrix} \pi_{1,k}^{m_k} \\ \vdots \\ \pi_{N,k}^{m_k} \end{pmatrix} = \begin{pmatrix} \gamma_k(m_k) \\ \vdots \\ \gamma_k(m_k + N - 1) \end{pmatrix}$$

for  $k = 1, 2$ . We will prove

$$\sup_N \sum_{i=1}^N |\pi_{i,k}^{m_k}| < \infty \quad (m_k \geq -N + 1) \quad (\text{A.2})$$

by the mathematical induction. In  $-N + 1 \leq m_k \leq 0$  ( $k = 1, 2$ ),  $\pi_{i,k}^{m_k} = \delta_{i,(-m_k+1)}$  from the uniqueness of  $(\pi_{1,k}^{m_k}, \pi_{2,k}^{m_k}, \dots, \pi_{N,k}^{m_k})'$  where  $\delta_{i,(-m_k+1)}$  is 1 if  $i = -m_k + 1$ , otherwise 0. Therefore, (A.2) is clear in this case. When  $m_k = 1$ ,

$$\tilde{\Sigma}_k \begin{pmatrix} \pi_{1,k}^1 \\ \vdots \\ \pi_{N,k}^1 \end{pmatrix} = \begin{pmatrix} \gamma_k(1) \\ \vdots \\ \gamma_k(N) \end{pmatrix}.$$

If  $z_t$  is  $\text{AR}(P_k)$  with the autocovariance function  $\gamma_k$ , that is  $z_t = \phi_{1,k}z_{t-1} + \dots + \phi_{P_k,k}z_{t-P_k} + \eta_t$  where  $\eta_t$  is a white noise,  $\pi_{j,k}^1 = \phi_{j,k}$  ( $j = 1, \dots, P_k$ ) and  $\pi_{j,k}^1 = 0$  ( $j > P_k$ ). Therefore,  $\sup_N \sum_{i=1}^N |\pi_{i,k}^1| < \infty$ . Next assume that (A.2) holds for  $m_k = m$ . Consider  $m_k = m + 1$ . Then, from a routine calculation,  $P_{\overline{\text{sp}}(z_1, \dots, z_N)} z_{N+m+1} = \sum_{l=1}^N \pi_{l,k}^{m+1} z_{N+1-l}$  where  $P_{\overline{\text{sp}}(z_1, \dots, z_N)}$  is the projection to the closed subspace  $\overline{\text{sp}}(z_1, \dots, z_N)$  spanned by  $(z_1, \dots, z_N)$ . Moreover, from the similar calculation,

$$\begin{aligned} P_{\overline{\text{sp}}(z_1, \dots, z_N)} z_{N+m+1} &= P_{\overline{\text{sp}}(z_1, \dots, z_N)} (P_{\overline{\text{sp}}(z_1, \dots, z_{N+1})} z_{N+m+1}) \\ &= P_{\overline{\text{sp}}(z_1, \dots, z_N)} \left( \sum_{l=1}^{N+1} \pi_{l,k}^m z_{N+2-l} \right) \\ &= P_{\overline{\text{sp}}(z_1, \dots, z_N)} \pi_{1,k}^m z_{N+1} + \sum_{l=2}^{N+1} \pi_{l,k}^m z_{N+2-l} \\ &= \pi_{1,k}^m \left( \sum_{l=1}^N \pi_{l,k}^1 z_{N+1-l} \right) + \sum_{l=2}^{N+1} \pi_{l,k}^m z_{N+2-l}. \end{aligned}$$

Therefore,  $\pi_{l,k}^{m+1} = \pi_{1,k}^m \pi_{l,k}^1 + \pi_{l+1,k}^m$  ( $l = 1, \dots, N$ ). In this case,

$$\sup_N \sum_{i=1}^N |\pi_{i,k}^{m+1}| \leq \sup_N \sum_{i=1}^N |\pi_{1,k}^m| |\pi_{i,k}^1| + \sup_N \sum_{i=1}^N |\pi_{i+1,k}^m| < \infty.$$

(A.2) holds for  $m_k = m + 1$ . Note that  $\pi_{i,j}^{m_1, m_2} = \pi_{i,1}^{m_1} \pi_{j,2}^{m_2}$  for  $1 \leq i, j \leq N$ . Then

$$\sum_{i=1}^N \sum_{j=1}^N |\pi_{i,j}^{m_1, m_2}| = \sum_{i=1}^N |\pi_{i,1}^{m_1}| \sum_{j=1}^N |\pi_{j,2}^{m_2}|$$

is bounded for  $N$ . □

**Lemma A.2.** Suppose that  $h(\lambda_1, \lambda_2)$  is a continuous function on  $[-\pi, \pi]^2$  and  $h(\lambda_1, \lambda_2) = h(-\lambda_1, \lambda_2) = h(\lambda_1, -\lambda_2) = h(-\lambda_1, -\lambda_2)$ . Then, for any sufficiently small  $\epsilon > 0$ , there exist  $h_L(\lambda_1, \lambda_2)$  and  $h_U(\lambda_1, \lambda_2)$  such that

$$\begin{aligned} h_L(\lambda_1, \lambda_2) &= \sum_{k_1=-K_1}^{K_1} \sum_{k_2=-K_2}^{K_2} a_{k_1, k_2} e^{i(k_1 \lambda_1 + k_2 \lambda_2)}, \quad a_{k_1, k_2} = a_{-k_1, -k_2}, \\ h_U(\lambda_1, \lambda_2) &= \sum_{k_1=-K_1}^{K_1} \sum_{k_2=-K_2}^{K_2} b_{k_1, k_2} e^{i(k_1 \lambda_1 + k_2 \lambda_2)}, \quad b_{k_1, k_2} = b_{-k_1, -k_2}, \\ h_L(\lambda_1, \lambda_2) &\leq h(\lambda_1, \lambda_2) \leq h_U(\lambda_1, \lambda_2) \end{aligned}$$

and

$$h_U(\lambda_1, \lambda_2) - h_L(\lambda_1, \lambda_2) \leq \epsilon, \quad (\lambda_1, \lambda_2) \in [-\pi, \pi]^2.$$

*Proof.* We set  $D_{n_1, n_2}(y_1, y_2) = \sum_{|j_1| \leq n_1} \sum_{|j_2| \leq n_2} e^{i(j_1 y_1 + j_2 y_2)} = D_{n_1}(y_1) D_{n_2}(y_2)$  where  $D_{n_i}(y_i) = \sum_{|j_i| \leq n_i} e^{ij_i y_i}$  ( $i = 1, 2$ ).  $D_{n_i}(y_i)$  is the Dirichlet kernel,

$$D_{n_i}(y_i) = \begin{cases} \frac{\sin[(n_i + \frac{1}{2})y_i]}{\sin(\frac{1}{2}y_i)} & \text{if } y_i \neq 0, \\ 2n_i + 1 & \text{if } y_i = 0. \end{cases}$$

Additionally, consider

$$\begin{aligned} S_{n_1, n_2} h(\lambda'_1, \lambda'_2) &= \sum_{|j_1| \leq n_1} \sum_{|j_2| \leq n_2} \langle h, e_{j_1, j_2} \rangle e_{j_1, j_2} \\ &= \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} h(\lambda_1, \lambda_2) \sum_{|j_1| \leq n_1} \sum_{|j_2| \leq n_2} e^{ij_1(\lambda'_1 - \lambda_1)} e^{ij_2(\lambda'_2 - \lambda_2)} d\lambda_1 d\lambda_2 \\ &= \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} h(\lambda_1, \lambda_2) D_{n_1}(\lambda'_1 - \lambda_1) D_{n_2}(\lambda'_2 - \lambda_2) d\lambda_1 d\lambda_2, \end{aligned}$$



where  $e_{j_1, j_2} = e^{i(j_1 \lambda_1 + j_2 \lambda_2)}$  and  $\langle h, e_{j_1, j_2} \rangle = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} h(\lambda_1, \lambda_2) e^{-i(j_1 \lambda_1 + j_2 \lambda_2)} d\lambda_1 d\lambda_2 / (2\pi)^2$ . By defining  $h(\lambda_1, \lambda_2) = h(\lambda_1 + 2\pi n'_1, \lambda_2 + 2\pi n'_2)$  ( $n'_1, n'_2 \in \mathbb{Z}$ ), it can be rewritten as

$$S_{n_1, n_2} h(\lambda'_1, \lambda'_2) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} h(\lambda'_1 - \lambda_1, \lambda'_2 - \lambda_2) D_{n_1}(\lambda_1) D_{n_2}(\lambda_2) d\lambda_1 d\lambda_2.$$

Moreover,

$$\frac{1}{n_1 n_2} \sum_{i_1=0}^{n_1-1} \sum_{i_2=0}^{n_2-1} S_{i_1, i_2} h(\lambda'_1, \lambda'_2) = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} h(\lambda'_1 - \lambda_1, \lambda'_2 - \lambda_2) K_{n_1}(\lambda_1) K_{n_2}(\lambda_2) d\lambda_1 d\lambda_2,$$

where  $K_{n_i}(y_i)$  is defined by

$$\frac{1}{2\pi n_i} \sum_{j_i=0}^{n_i-1} D_{j_i}(y_i) \quad (i = 1, 2)$$

and is called the Fejer kernel. Finally, it follows from the argument extended from Theorem 2.11.1 of Brockwell and Davis (1991; page 69) and the axial symmetry of  $h(\lambda_1, \lambda_2)$  that for any  $\epsilon > 0$ ,

$$\left| h(\lambda_1, \lambda_2) - \frac{1}{n_1 n_2} \sum_{i_1=0}^{n_1-1} \sum_{i_2=0}^{n_2-1} S_{i_1, i_2} h(\lambda_1, \lambda_2) \right| < \epsilon,$$

uniformly on  $[-\pi, \pi]^2$  for sufficiently large  $n_1$  and  $n_2$ . In that case,

$$\frac{1}{n_1 n_2} \sum_{i_1=0}^{n_1-1} \sum_{i_2=0}^{n_2-1} S_{i_1, i_2} h(\lambda_1, \lambda_2) = \sum_{j_1=-(n_1-1)}^{n_1-1} \sum_{j_2=-(n_2-1)}^{n_2-1} c_{j_1, j_2} e_{j_1, j_2},$$

where  $c_{j_1, j_2} = c_{-j_1, j_2} = c_{j_1, -j_2} = c_{-j_1, -j_2}$  from the axial symmetry of  $h(\lambda_1, \lambda_2)$ . Therefore,  $c_{j_1, j_2}$ 's are real and

$$\sum_{j_1=-(n_1-1)}^{n_1-1} \sum_{j_2=-(n_2-1)}^{n_2-1} c_{j_1, j_2} e_{j_1, j_2} = \sum_{j_1=-(n_1-1)}^{n_1-1} \sum_{j_2=-(n_2-1)}^{n_2-1} c_{j_1, j_2} (\cos(j_1 \lambda_1 + j_2 \lambda_2) + i \sin(j_1 \lambda_1 + j_2 \lambda_2)).$$

Now it follows that

$$\sum_{j_1=-(n_1-1)}^{n_1-1} \sum_{j_2=-(n_2-1)}^{n_2-1} c_{j_1, j_2} \sin(j_1 \lambda_1 + j_2 \lambda_2) = 0.$$

Therefore,  $\sum_{|j_1| \leq n_1-1} \sum_{|j_2| \leq n_2-1} c_{j_1, j_2} e_{j_1, j_2}$  is real. For any  $\epsilon > 0$ ,

$$\sum_{j_1=-(n_1-1)}^{n_1-1} \sum_{j_2=-(n_2-1)}^{n_2-1} c_{j_1, j_2} e_{j_1, j_2} - \frac{\epsilon}{2} \leq h(\lambda_1, \lambda_2) \leq \sum_{j_1=-(n_1-1)}^{n_1-1} \sum_{j_2=-(n_2-1)}^{n_2-1} c_{j_1, j_2} e_{j_1, j_2} + \frac{\epsilon}{2}$$

for sufficiently large  $n_1$  and  $n_2$ . Therefore, setting  $K_1 = n_1 - 1, K_2 = n_2 - 1, a_{k_1, k_2} = b_{k_1, k_2} = c_{k_1, k_2}$  except  $a_{0,0} = c_{0,0} - \epsilon/2, b_{0,0} = c_{0,0} + \epsilon/2$ , the proof is completed.  $\square$

## Appendix B : Proof of Theorem 4

We prove Theorem 4 by the arguments extended from Anderson (1971) and Yajima (1994).

### Proof of Theorem 4

$$\begin{aligned} & D_{N^2} E \left[ (\hat{\beta}_{PBE} - \beta)(\hat{\beta}_{PBE} - \beta)' \right] D_{N^2} \\ &= \left( D_{N^2}^{-1} X' \tilde{\Sigma}^{-1} X D_{N^2}^{-1} \right)^{-1} \left( D_{N^2}^{-1} X' \tilde{\Sigma}^{-1} \Sigma \tilde{\Sigma}^{-1} X D_{N^2}^{-1} \right) \left( D_{N^2}^{-1} X' \tilde{\Sigma}^{-1} X D_{N^2}^{-1} \right)^{-1}. \end{aligned}$$

From Theorem 1 the first and third terms converge to

$$\left( D_{N^2}^{-1} X' \tilde{\Sigma}^{-1} X D_{N^2}^{-1} \right)^{-1} \rightarrow \left( \frac{1}{(2\pi)^2} \int_{\Pi^2} \frac{1}{g(\lambda_1, \lambda_2)} dM(\lambda_1, \lambda_2) \right)^{-1}$$

as  $N \rightarrow \infty$ . Here Put

$$\begin{aligned} f(\lambda_1, \lambda_2) &= \frac{f(\lambda_1, \lambda_2)}{g(\lambda_1, \lambda_2)} \times g(\lambda_1, \lambda_2) \\ &= h(\lambda_1, \lambda_2) \times g(\lambda_1, \lambda_2). \end{aligned}$$

Firstly consider the case of

$$h(\lambda_1, \lambda_2) = \sum_{l_1=-L_1}^{L_1} \sum_{l_2=-L_2}^{L_2} b_{l_1, l_2} e^{i(l_1 \lambda_1 + l_2 \lambda_2)}, \quad b_{l_1, l_2} = b_{-l_1, -l_2}.$$

In this case,

$$\begin{aligned} E[\epsilon_{(j_1, j_2)} \epsilon_{(k_1, k_2)}] &= \gamma_\epsilon(j_1 - k_1, j_2 - k_2) \\ &= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{i((j_1 - k_1)\lambda_1 + (j_2 - k_2)\lambda_2)} f(\lambda_1, \lambda_2) d\lambda_1 d\lambda_2 \\ &= \sum_{l_1=-L_1}^{L_1} \sum_{l_2=-L_2}^{L_2} b_{l_1, l_2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{i((j_1 - k_1 + l_1)\lambda_1 + (j_2 - k_2 + l_2)\lambda_2)} g(\lambda_1, \lambda_2) d\lambda_1 d\lambda_2. \end{aligned}$$

Therefore,

$$\Sigma = \sum_{l_1=-L_1}^{L_1} \sum_{l_2=-L_2}^{L_2} b_{l_1, l_2} \tilde{\Sigma}(l_1, l_2),$$

where

$$\tilde{\Sigma}(l_1, l_2) = \begin{pmatrix} \gamma(l_1, l_2) & \gamma(l_1, l_2 - 1) & \cdots & \gamma(l_1 + 1 - N, l_2) \\ \gamma(l_1, l_2 + 1) & \gamma(l_1, l_2) & \cdots & \gamma(l_1 + 1 - N, l_2 + 1) \\ \vdots & \vdots & & \vdots \\ \gamma(l_1 + N - 1, l_2) & \gamma(l_1 + N - 1, l_2 - 1) & \cdots & \gamma(l_1, l_2) \\ \vdots & \vdots & & \vdots \\ \gamma(l_1 + N - 1, l_2 + N - 1) & \gamma(l_1 + N - 1, l_2 + N - 2) & \cdots & \gamma(l_1, l_2 + N - 1) \\ \cdots & \gamma(l_1 + 1 - N, l_2 + 2 - N) & \gamma(l_1 + 1 - N, l_2 + 1 - N) & \\ \cdots & \gamma(l_1 + 1 - N, l_2 + 3 - N) & \gamma(l_1 + 1 - N, l_2 + 2 - N) & \\ & \vdots & \vdots & \\ \cdots & \gamma(l_1, l_2 + 2 - N) & \gamma(l_1, l_2 + 1 - N) & \\ & \vdots & \vdots & \\ \cdots & \gamma(l_1, l_2 + 1) & \gamma(l_1, l_2) & \end{pmatrix}.$$

Consider the case of  $l_1 \geq 0$  and  $l_2 \geq 0$  for sufficiently large  $N$ . In that case,

the  $(m, n)$ th element  $\left(D_{N^2}^{-1} X' \tilde{\Sigma}^{-1} \tilde{\Sigma}(l_1, l_2) \tilde{\Sigma}^{-1} X D_{N^2}^{-1}\right)_{m,n}$  is

$$\begin{aligned}
& \frac{1}{\left(a_{mm}^{(N,N)}(0,0)a_{nn}^{(N,N)}(0,0)\right)^{1/2}} \left\{ (\mathbf{x}'_{(1,\cdot),m}, \dots, \mathbf{x}'_{(N,\cdot),m}) \tilde{\Sigma}^{-1} \right. \\
& \times \left( x_{(l_1+1,l_2+1),n}, \dots, x_{(l_1+1,N),n}, 0, \dots, 0, \dots, x_{(N,l_2+1),n}, \dots, x_{(N,N),n}, 0, \dots, 0, \right. \\
& \left. 0, \dots, 0 \right)' + (\mathbf{x}'_{(1,\cdot),m}, \dots, \mathbf{x}'_{(N,\cdot),m}) \tilde{\Sigma}^{-1} \\
& \times \left( 0, \dots, 0, \sum_{i=1}^N \sum_{j=1}^N \pi_{i,j}^{l_1+1-N,1} x_{(N+1-i,N+1-j),n}, \dots, \sum_{i=1}^N \sum_{j=1}^N \pi_{i,j}^{l_1+1-N,l_2} \right. \\
& \times x_{(N+1-i,N+1-j),n}, \dots, 0, \dots, 0, \sum_{i=1}^N \sum_{j=1}^N \pi_{i,j}^{0,1} x_{(N+1-i,N+1-j),n}, \dots, \sum_{i=1}^N \sum_{j=1}^N \pi_{i,j}^{0,l_2} \\
& \times x_{(N+1-i,N+1-j),n}, \sum_{i=1}^N \sum_{j=1}^N \pi_{i,j}^{1,l_2+1-N} x_{(N+1-i,N+1-j),n}, \dots, \\
& \left. \left. \sum_{i=1}^N \sum_{j=1}^N \pi_{i,j}^{l_1,l_2} x_{(N+1-i,N+1-j),n} \right)' \right\} \quad (\text{B.1})
\end{aligned}$$

where  $\mathbf{x}_{(i,\cdot),m} = (x_{(i,1),m}, \dots, x_{(i,N),m})'$  ( $i = 1, \dots, N$ ) and  $\{\pi_{i,j}^{m_1,m_2}\}$  is defined in Lemma A. 1. It follows from the property of Kronecker product (see Horn and Johnson 1991), the exact form of the inverse matrix of  $\text{AR}(P_i)$  ( $i = 1, 2$ ), (a)-(d) and a routine calculation that the first term of (B.1) converges to

$$\frac{1}{(2\pi)^2} \int_{\Pi^2} \frac{e^{-i(l_1\lambda_1 + l_2\lambda_2)}}{g_1(\lambda_1)g_2(\lambda_2)} dM_{mn}(\lambda_1, \lambda_2)$$

as  $N \rightarrow \infty$ . Next

$$\begin{aligned}
& |\text{the second term of (B.1)}| \leq \frac{1}{\left(a_{mm}^{(N,N)}(0,0)a_{nn}^{(N,N)}(0,0)\right)^{1/2}} \\
& \times \left\{ (\mathbf{x}'_{(1,\cdot),m}, \dots, \mathbf{x}'_{(N,\cdot),m}) \tilde{\Sigma}^{-1} (\mathbf{x}'_{(1,\cdot),m}, \dots, \mathbf{x}'_{(N,\cdot),m})' \right\}^{1/2} \\
& \times \left\{ \left( 0, \dots, \sum_{i=1}^N \sum_{j=1}^N \pi_{i,j}^{l_1,l_2} x_{(N+1-i,N+1-j),n} \right) \tilde{\Sigma}^{-1} \left( 0, \dots, \sum_{i=1}^N \sum_{j=1}^N \pi_{i,j}^{l_1,l_2} x_{(N+1-i,N+1-j),n} \right)' \right\}^{1/2}
\end{aligned}$$

$$\begin{aligned}
&\leq \frac{1}{\left(a_{mm}^{(N,N)}(0,0)a_{nn}^{(N,N)}(0,0)\right)^{1/2}} \left(\tilde{\lambda}_{N^2} \sum_{i=1}^N \sum_{j=1}^N x_{(i,j),m}^2\right)^{1/2} \\
&\times \left\{ \tilde{\lambda}_{N^2} \left( \sum_{m_1=l_1+1-N}^0 \sum_{m_2=1}^{l_2} \left( \sum_{i=1}^N \sum_{j=1}^N \pi_{i,j}^{m_1,m_2} x_{(N+1-i,N+1-j),n} \right)^2 \right. \right. \\
&\left. \left. + \sum_{m_1=1}^{l_1} \sum_{m_2=l_2+1-N}^{l_2} \left( \sum_{i=1}^N \sum_{j=1}^N \pi_{i,j}^{m_1,m_2} x_{(N+1-i,N+1-j),n} \right)^2 \right) \right\}^{1/2} \\
&\leq \tilde{\lambda}_{N^2} \left( \frac{N \max_{1 \leq t_1, t_2 \leq N} x_{(t_1, t_2), n}^2}{a_{nn}^{(N,N)}(0,0)} \right)^{1/2} \left\{ \sum_{m_1=1}^{l_1} \left( \sum_{i=1}^N |\pi_{i,1}^{m_1}| \right)^2 \right. \\
&\left. + \sum_{m_2=1}^{l_2} \left( \sum_{j=1}^N |\pi_{j,2}^{m_2}| \right)^2 + \sum_{m_1=1}^{l_1} \sum_{m_2=1}^{l_2} \left( \sum_{i=1}^N \sum_{j=1}^N |\pi_{i,j}^{m_1,m_2}| \right)^2 \right\}, \quad (\text{B.2})
\end{aligned}$$

where  $\tilde{\lambda}_{N^2}$  is the greatest eigenvalue of  $\tilde{\Sigma}^{-1}$  and  $(\pi_{1,k}^{m_k}, \dots, \pi_{N,k}^{m_k})'$  ( $k = 1, 2$ ) is defined in the proof of Lemma A.1. Since it is proved that  $\tilde{\lambda}_{N^2}$  is bounded from the property of Kronecker product (see Horn and Johnson 1991) and Proposition 4.5.3 of Brockwell and Davis (1991), (B.2) converges to 0 as  $N \rightarrow \infty$  from (g), Lemma A.1 and its proof. Therefore,

$$\left( D_{N^2}^{-1} X' \tilde{\Sigma}^{-1} \tilde{\Sigma}(l_1, l_2) \tilde{\Sigma}^{-1} X D_{N^2}^{-1} \right)_{m,n} \rightarrow \frac{1}{(2\pi)^2} \int_{\Pi^2} \frac{e^{-i(l_1 \lambda_1 + l_2 \lambda_2)}}{g_1(\lambda_1) g_2(\lambda_2)} dM_{mn}(\lambda_1, \lambda_2)$$

as  $N \rightarrow \infty$ . In a similar way, we can show other cases of  $(l_1 \geq 0, l_2 < 0)$ ,  $(l_1 < 0, l_2 \geq 0)$  and  $(l_1 < 0, l_2 < 0)$ . Then

$$\begin{aligned}
\left( D_{N^2}^{-1} X' \tilde{\Sigma}^{-1} \Sigma \tilde{\Sigma}^{-1} X D_{N^2}^{-1} \right)_{m,n} &= \sum_{l_1=-L_1}^{L_1} \sum_{l_2=-L_2}^{L_2} b_{l_1, l_2} \left( D_{N^2}^{-1} X' \tilde{\Sigma}^{-1} \tilde{\Sigma}(l_1, l_2) \tilde{\Sigma}^{-1} X D_{N^2}^{-1} \right)_{m,n} \\
&\rightarrow \frac{1}{(2\pi)^2} \int_{\Pi^2} \frac{f(\lambda_1, \lambda_2)}{g^2(\lambda_1, \lambda_2)} dM_{mn}(\lambda_1, \lambda_2)
\end{aligned}$$

as  $N \rightarrow \infty$ . Finally consider the case of general  $h(\lambda_1, \lambda_2)$ . From (f), (h) and Lemma A.2, for any sufficiently small  $\epsilon > 0$ , there exist  $h_L(\lambda_1, \lambda_2)$  and

$h_U(\lambda_1, \lambda_2)$  such that

$$\begin{aligned} h_L(\lambda_1, \lambda_2) &= \sum_{k_1=-K_1}^{K_1} \sum_{k_2=-K_2}^{K_2} a_{k_1, k_2} e^{i(k_1 \lambda_1 + k_2 \lambda_2)}, \quad a_{k_1, k_2} = a_{-k_1, -k_2}, \\ h_U(\lambda_1, \lambda_2) &= \sum_{k_1=-K_1}^{K_1} \sum_{k_2=-K_2}^{K_2} b_{k_1, k_2} e^{i(k_1 \lambda_1 + k_2 \lambda_2)}, \quad b_{k_1, k_2} = b_{-k_1, -k_2}, \\ h_L(\lambda_1, \lambda_2) &\leq h(\lambda_1, \lambda_2) \leq h_U(\lambda_1, \lambda_2) \end{aligned}$$

and

$$h_U(\lambda_1, \lambda_2) - h_L(\lambda_1, \lambda_2) \leq \epsilon, \quad (\lambda_1, \lambda_2) \in [-\pi, \pi]^2.$$

Then for any  $\gamma \in \mathbb{R}^p$ ,

$$\gamma' D_{N^2}^{-1} X' \tilde{\Sigma}^{-1} \Sigma_L \tilde{\Sigma}^{-1} X D_{N^2}^{-1} \gamma \leq \gamma' D_{N^2}^{-1} X' \tilde{\Sigma}^{-1} \Sigma \tilde{\Sigma}^{-1} X D_{N^2}^{-1} \gamma \leq \gamma' D_{N^2}^{-1} X' \tilde{\Sigma}^{-1} \Sigma_U \tilde{\Sigma}^{-1} X D_{N^2}^{-1} \gamma,$$

where  $\Sigma_L$  and  $\Sigma_U$  are the covariance matrices with the spectral density functions

$h_L(\lambda_1, \lambda_2)g(\lambda_1, \lambda_2)$  and  $h_U(\lambda_1, \lambda_2)g(\lambda_1, \lambda_2)$  respectively. From the above discussion,

$$\begin{aligned} \gamma' \frac{1}{(2\pi)^2} \int_{\Pi^2} \frac{h_L(\lambda_1, \lambda_2)}{g(\lambda_1, \lambda_2)} dM(\lambda_1, \lambda_2) \gamma &\leq \lim_{N \rightarrow \infty} \gamma' D_{N^2}^{-1} X' \tilde{\Sigma}^{-1} \Sigma \tilde{\Sigma}^{-1} X D_{N^2}^{-1} \gamma \\ &\leq \overline{\lim}_{N \rightarrow \infty} \gamma' D_{N^2}^{-1} X' \tilde{\Sigma}^{-1} \Sigma \tilde{\Sigma}^{-1} X D_{N^2}^{-1} \gamma \leq \gamma' \frac{1}{(2\pi)^2} \int_{\Pi^2} \frac{h_U(\lambda_1, \lambda_2)}{g(\lambda_1, \lambda_2)} dM(\lambda_1, \lambda_2) \gamma. \end{aligned}$$

Since  $\epsilon$  is arbitrary, as  $N \rightarrow \infty$

$$\begin{aligned} \gamma' D_{N^2}^{-1} X' \tilde{\Sigma}^{-1} \Sigma \tilde{\Sigma}^{-1} X D_{N^2}^{-1} \gamma &\rightarrow \gamma' \frac{1}{(2\pi)^2} \int_{\Pi^2} \frac{h(\lambda_1, \lambda_2)}{g(\lambda_1, \lambda_2)} dM(\lambda_1, \lambda_2) \gamma \\ &= \gamma' \frac{1}{(2\pi)^2} \int_{\Pi^2} \frac{f(\lambda_1, \lambda_2)}{g^2(\lambda_1, \lambda_2)} dM(\lambda_1, \lambda_2) \gamma. \end{aligned}$$

Since this holds for every vector  $\gamma$ , as  $N \rightarrow \infty$

$$D_{N^2}^{-1} X' \tilde{\Sigma}^{-1} \Sigma \tilde{\Sigma}^{-1} X D_{N^2}^{-1} \rightarrow \frac{1}{(2\pi)^2} \int_{\Pi^2} \frac{f(\lambda_1, \lambda_2)}{g^2(\lambda_1, \lambda_2)} dM(\lambda_1, \lambda_2).$$

The proof is completed. □

**Acknowledgements** The author is grateful to Professor Yoshihiro Yajima for helpful comments and discussions.

## References

- [1] Amemiya, T. (1973). Generalized least squares with an estimated autocovariance matrix. *Econometrica*, **41**, 723-732.
- [2] Anderson, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.
- [3] Banerjee, S., Gelfand, A. E., Finley, A. O. and Sang, H. (2008). Gaussian predictive process models for large spatial datasets. *Journal of the Royal Statistical Society: Series B*, **70**, 825-848.
- [4] Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods*, second ed. Springer, New York.
- [5] Cohen, G. and Francos, J. M. (2002). Linear least squares estimation of regression models for two-dimensional random fields. *Journal of Multivariate Analysis*, **82**, 431-444.
- [6] Cressie, N. (1993). *Statistics for Spatial Data*, revised ed. Wiley, New York.
- [7] Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B*, **70**, 209-226.
- [8] Engle, R.F. (1974). Specification of the disturbance for efficient estimation. *Econometrica*, **42**, 135-146.
- [9] Fuentes, M. (2005). A formal test for nonstationarity of spatial stochastic processes. *Journal of Multivariate Analysis*, **96**, 30-54.
- [10] Furrer, R., Genton, M. G. and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, **15**, 502-523. *Erratum and Addendum : Journal of Computational and Graphical Statistics*, **21**, 823-824.
- [11] Genton, M. G. (2007). Separable approximations of space-time covariance matrices. *Environmetrics*, **18**, 681-695.
- [12] Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series*. Wiley, New York.
- [13] Horn, R. A. and Johnson, C. R. (1991). *Topics in Matrix Analysis*. Cambridge University Press.

- [14] Huang, H.-C., Hsu, N.-J., Theobald, D.M. and Breidt, F.J. (2010). Spatial lasso with applications to GIS model selection. *Journal of Computational and Graphical Statistics*, **19**, 963-983.
- [15] Isaaks, E. H. and Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics*. Oxford University Press.
- [16] Koreisha, S. G. and Fang, Y. (2001). Generalized least squares with misspecified serial correlation structures. *Journal of the Royal Statistical Society: Series B*, **63**, 515-531.
- [17] Lindgren, F., Rue, H. and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The SPDE approach (with discussion). *Journal of the Royal Statistical Society: Series B*, **73**, 423-498.
- [18] Taniguchi, M., Hirukawa, J. and Tamaki, K. (2008). *Optimal Statistical Inference in Financial Engineering*. Financial Mathematics Series. Chapman & Hall/CRC, New York.
- [19] Toyooka, Y. (1985). Second-order risk comparison of SLSE with GLSE and MLE in a regression with serial correlation. *Journal of Multivariate Analysis*, **17**, 107-126.
- [20] Yajima, Y. (1994). On semiparametric estimation of time series analysis and its applications. *The Journal of economics*, **59(4)**, 2-22 (in Japanese).
- [21] Yajima, Y. and Matsuda, Y. (2008). Asymptotic properties of the LSE of a spatial regression in both weakly and strongly dependent stationary random fields. Discussion Paper CIRJE-F-587, Graduate School of Economics, University of Tokyo.



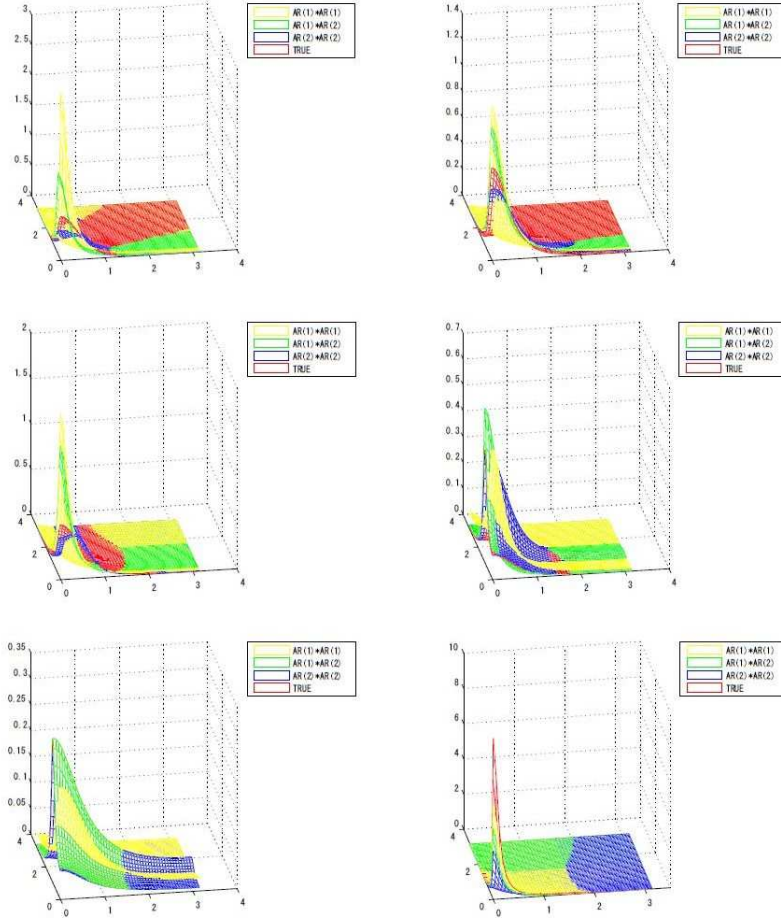


Figure 1: Approximated spectral density functions on  $[0, \pi]^2$ . The true covariance function corresponds to the isotropic Matérn ( $\nu = 2.0$ ), the isotropic Matérn ( $\nu = 1.0$ ) and  $(\nu = 2.0) \times (\nu = 1.0)$  for the left, middle and right column in the top row respectively. Similarly, it corresponds to  $(\nu = 1.0) \times AR(2)$ ,  $AR(1) \times AR(2)$  and  $AR(1) \times AR(1)$  for the left, middle and right column in the bottom row respectively.